Images as Evidence: Forensic Examination of Scientific Images[1]

John W. Krueger, Division of Investigative Oversight, Office of Research Integrity, DHHS, USA

Keywords: Autoradiograms, Blots, Image processing, Manipulation and falsification, Scientific misconduct

A “questioned” scientific image, i.e., suspicions of falsification (or plagiarism) of image data, such as photographs of PAGE gels, autoradiograms, and blots (Western, Northern, and Southern) can give rise to an allegation of misconduct in science. Pursuing oversight review of institutional investigations and reviewing allegations that ORI receives directly, ORI commonly examines the evidence through image processing. Typically, the examination can extend beyond merely asking “what is the evidence the image is/isn’t authentic?” and/or “are two contested images really the same?” Examples from these cases illustrate the general principles in forensic image processing and several methods that ORI has found useful in resolving the questions at hand. They provide an opportunity for further instruction as to what constitutes data falsification in an image.

Design/Methods

Source of Material: The material for this presentation was taken from a survey of 19 ORI cases that involved allegations of falsification or plagiarism of the images of gels, blots, auto-radiograms, and micrographs. The cases span a period from 1990 to 2000. The number of such questioned image allegations has generally increased, as has their incidence relative to other ORI cases. (Figure 9) A compilation from this review is discussed below.

Software: Most of ORI’s image analysis was done on a Macintosh® computer. The reason is both historical and practical; files transfer easily from the Windows® platform to the Macintosh®; but the opposite is not always true.

ORI has found several different image processing programs that are readily available and well documented so that the results can be easily shared with all parties in a potentially adversarial dispute. (1, 2) Each separately —or in combination with the others— offers distinct advantages. The image processing was conducted using either NIH Image (3) and/or Adobe Photoshop® (4), both of which were equipped with the Image Processing Tool Kit® (IPTK) plugins. (5) NIH Image, developed at the National Institutes of Health, is in the public domain and is ideal for analytical treatment of 8 bit (256 shades) monochromatic images. Photoshop is better suited for conducting overlay comparisons of two images and for working with color, but it requires the IPTK’s plugins for analytical work. Finally, ImageJ (6) is an update of the NIH public domain software that is compatible across computer platforms and will process images at 8, 16, and 32 bit depth; thus, it can detect vastly fainter features that might be hidden in the image.

Other Resources: Articles that can serve as guidance to issues involved in the forensic examination of contested documents can be obtained on the Internet (1, 2). Those sites can serve as

Corresponding author: John Krueger, Office of Research Integrity, 5515 Security Lane, Suite 700, Rockville, MD, 20852, 301-443-2263 (voice), 301-594-0039 (fax), jkrueger@osophs.dhhs.gov.

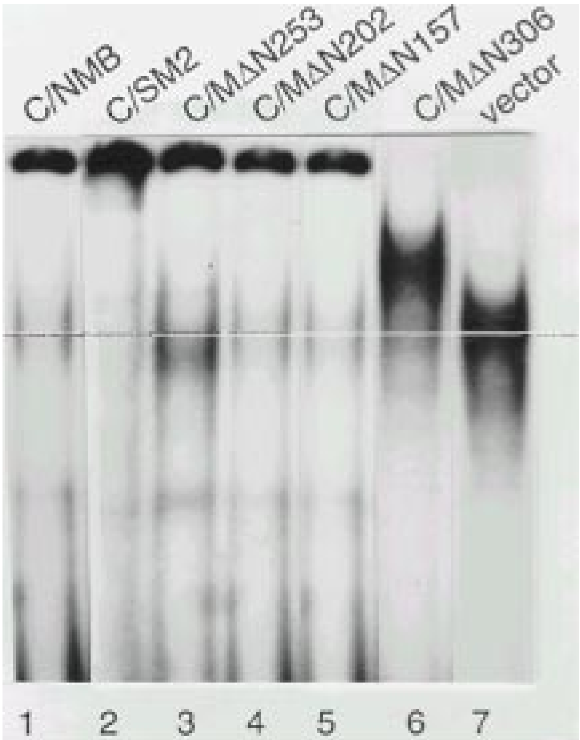

Figure 1. Original Western blot data. The results of an electrophoretic mobility shift assay to show bands reflecting the gene expression of separate mutant proteins. However, the shape of the bands and the pattern of the background in the 1^{st}, 4^{th}, and 5^{th} lanes look alike.

links to find other material.

Reasons for Examination and Some Principles of the Image Analysis Methods

The usual motivation for image analysis is to examine the authenticity of a particular document or to determine whether two purportedly different images really were derived from the same experiment.[2] In fact, image analysis provides information that addresses other issues. For example, features can be detected that reveal the source of the image, whether it is compatible with laboratory records such as autoradiograms or prior blots (see note 2), and whether the questioned image existed on a computer as a file, or on a website as a component of someone else's homepage. Second, the analysis of the latter sources can provide dates of creation, which can be corroborated with laboratory records, etc. Third, image enhancement may reveal evidence for the mechanics of the figure's construction, such as edges of photographic prints and presence of "white-out" and may uncover "hidden" details, such as erasures of labels. Fourth, an analysis of the new facts produced, such as the sources, dates, and incidence of re-use, may establish whether a pattern of misrepresentation existed that rules out honest error. Examples from ORI's cases illustrate these points.

Figure 1 represents a photographic mock-up of Western blot data, consisting of five photographic strips, in which the 2^{nd} to 4^{th} lanes were on one strip. Although purportedly showing differentdeterminations of protein created by separate mutant gene constructs, the 1^{st}, 4^{th}, and 5^{th} lanes look unexpectedly similar, but it is difficult to say for certain that they are the same.

One generic principle in making comparisons to determine the authenticity of data is to look at the features that would otherwise be un-noteworthy, such as fine features hidden in the background.[3] There may be random features that are hidden from our perception. The human eye, which responds to contrast, can distinguish only ~50 shades of gray (7) or less (8), but it can detect 100 shades of color (8).[4] However, the computer's response is not dependant on contrast; it can selectively amplify very slight differences in shade. The ability to detect such differences can be affected by the "depth" used to digitize the image, which in this case is 256 shades of gray.[5] The amplified differences in gray shades can next be shadowed and assigned false-colors to make faint differences even more visible, as shown in Figure 2.

These steps reveal faint artifactual features that were "hidden" in the background which are common to three of the lanes. Thus the respondent's claim, that at least two of the three lanes (1, 4, or 5 in Figure 1) represented evidence for gene expression of different mutant proteins, was a clear falsification of data.

Enhancement of the small difference in shades can also expose minute structural details in the morphology of bands, which otherwise would look smooth and featureless. Figure 3 illustrates a photo-montage from the above case; the bands appear similar in the 1^{st} and 5^{th} lanes.

Contrast enhancement and false-coloring of the above image as shown in Figure 4 demonstrate that the respective bands share similar miniature features. Thus, the image analysis showed that the first and the last lanes were from the same experiment.

In both examples above, the falsification was associated with false labeling of data that had been "re-used" from another experiment. The

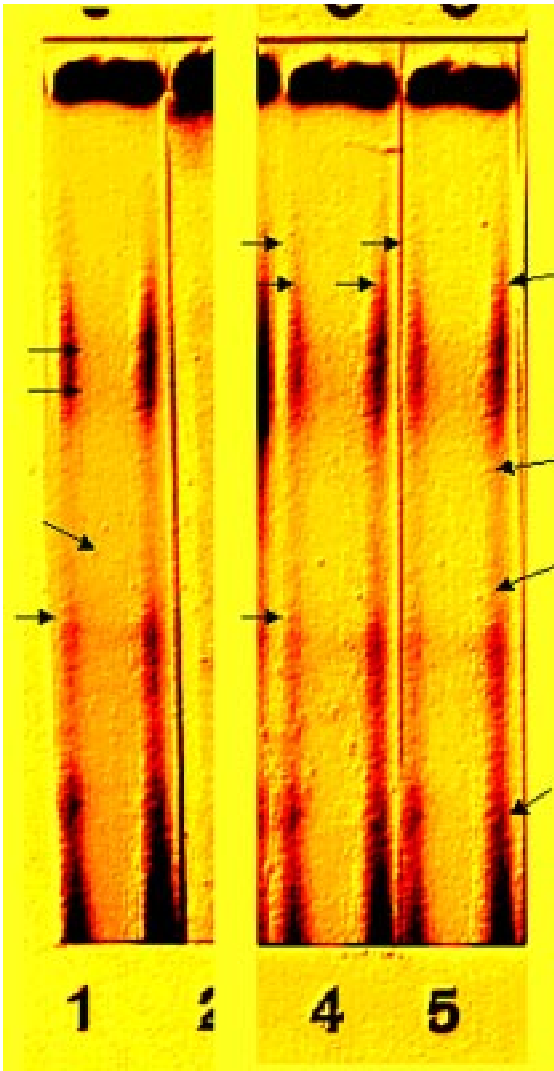

Figure 2 (left). Image enhancement of the questioned Western blot data. This ORI image analysis figure shows only the 1st, 4th, and 5th lanes from Figure 1. Contrast enhancement of the monochromatic gray-scale image, followed by shadowing and false-coloring (using NIH Image), revealed small features in the background artifact that are common to all three lanes (arrows) which the respondent had falsely represented as different. Note that in this case some differences can also be found, such as an horizontal artifact under the top band in the 4th lane, but they are in the background and represent artifacts that were introduced at some later point.

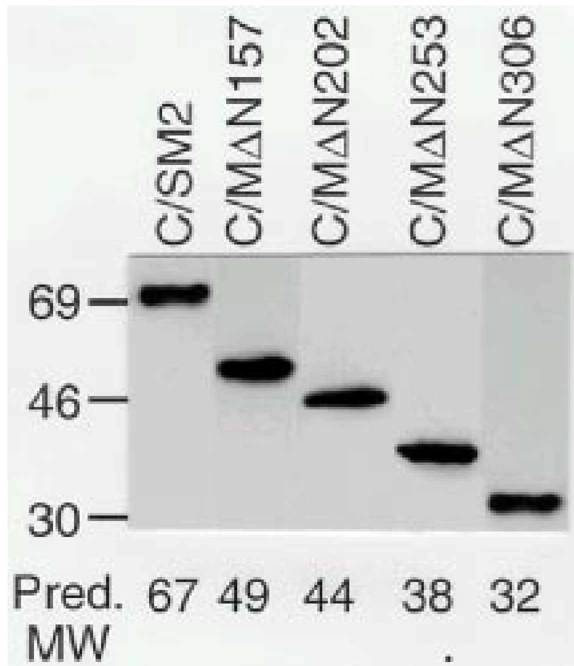

Figure 3. Western blot data. The results purportedly found a good agreement between the observed and the predicted size of five mutant proteins. However, the 1st and the 5th lanes' bands look similar.

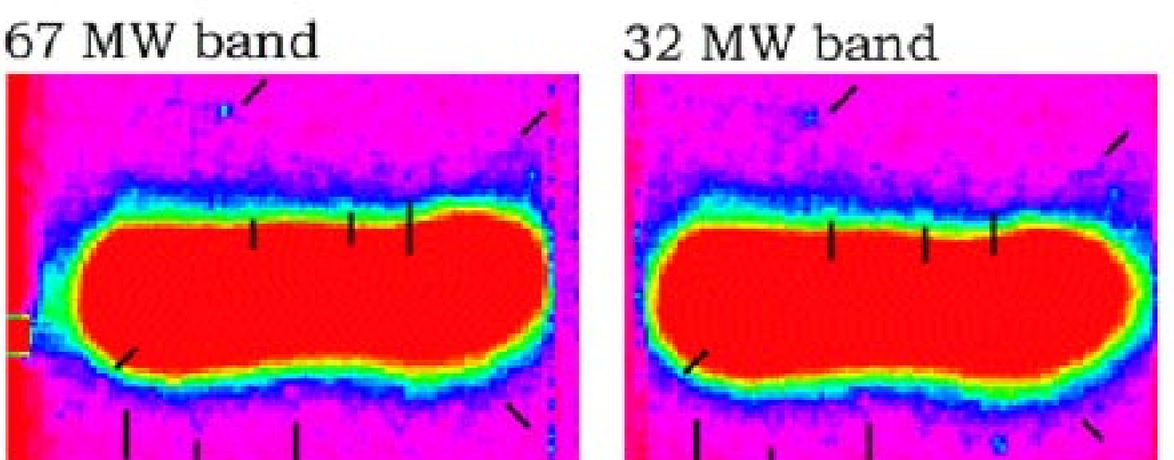

Figure 4. Image enhancement of the 67 kDa MW and 32 kDa MW bands from Figure 3. The bold lines denote miniature features in the bands' morphology that indicate both were actually the same data, which the respondent had falsified in re-use.

second example showed an additional falsification involving a false claim that the molecular weights had been determined. In this case, the intent to falsify the data is *prima facie*, because the molecular weight could not have been measured for the last re-used band. Finally, because the molecular weights were purported to approach the predicted values, the evidence also indicates that the falsifications are significant. These elements strengthen the findings.

Background detail and miniature features cannot be examined by image enhancement in all specimens. Fortunately, numerous other approaches are available in image processing to compare two questioned images. In general a combination of methods is determinative. For example, the morphology, size, and vertical arrangement of the bands and the existence of larger artifacts are the most obvious features to compare. Moreover, the horizontal spacing between the bands should not be overlooked; because substances rarely migrate on gels absolutely parallel, there may be slight differences in lateral disposition that are also significant. Some forms of artifact might re-occur, such as that introduced by a faulty film dryer and/or the edge of a blot on an autoradiographic film. The key question in cases of "replicating" artifacts is whether a relationship to other features should exist.

How to best visually represent the results of an image overlay is always a challenge. A visually effective and efficient method is to overlap color-coded plots of the "contour" map of the intensities in two separate blots, where the areas of overlap generate a third color. If two gray scale images are overlaid, the interpretation of the origin of features in the overlay becomes problematic unless each is first converted to a suitably chosen monochrome color scheme.

<u>Reconstruction of a Missing Document</u>: Analysis of an image can also be used to test the proffered source of a questioned image under circumstances in which the original raw data are missing. Figure 5 represents a composite image, which was created by combining a figure of a questioned Northern blot in the mentor's manuscript with a figure of a different experiment shown in the student's thesis. Unfortunately, the original blot and its PhosphoImager computer file were missing, but the mentor provided laboratory data purporting to be a different representation of the same blot (an ethidium bromide stain) that showed two groups

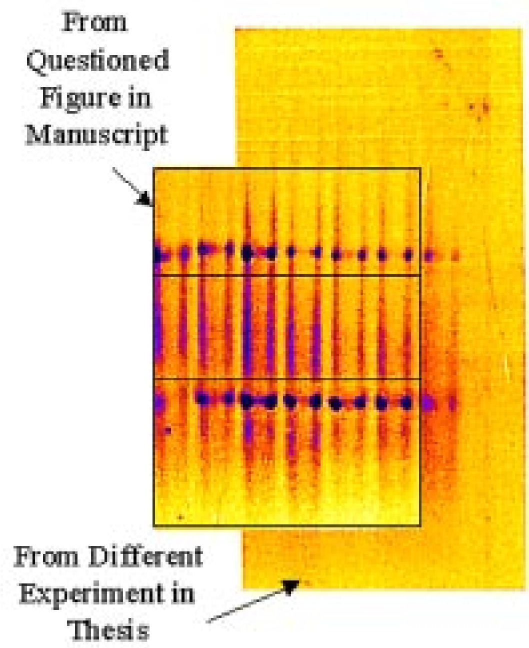

Figure 5. Overlay of the mentor's Northern blot mRNA data (small rectangle) with a figure from a different experiment from the student's thesis (tall rectangle). In this ORI image analysis, the actual fit was determined mathematically and showed the missing blot actually had at least *seven* lanes, indicating the respondent's claim was false.

of six lanes, separate by an empty lane. However, the overlay, shown in Figure 5, which was established as the best mathematical fit between the two sources, demonstrated that the missing original blot had to have had at least seven lanes. Thus, the proffered laboratory records could not be evidence of the mentor's "missing" data.

<u>Analysis of Poor Images</u>: The poor quality of an image is not necessarily a deterrent to the application of the above tools to its examination. The left side of Figure 6 shows a poor quality photocopy of data that was submitted in a mentor's National Institutes of Health (NIH) grant application, which purported to be a Western blot of an immunologic protein, "P-48," using ^{125}I-labeled human lymphocytes. The figure on the right side of Figure 6 represents the enhanced image of an autoradiogram from his student's experiments, which used ^{35}S - methionine labeling of cultured rats cells.

The distracting artifact due to over-photocopying could be minimized by image processing. This step revealed additional bands in the former with more certainty, and it more

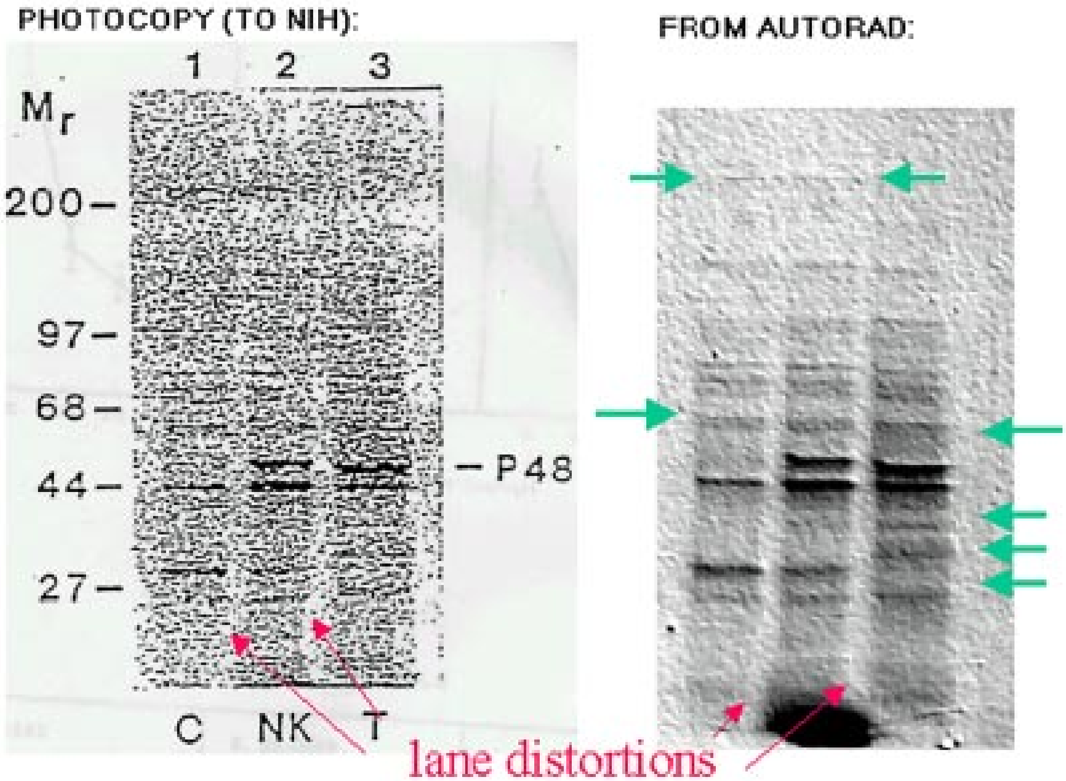

Figure 6. Examination of a poor quality photocopy. The mentor submitted the left hand "^{125}I-labeled" figure in an NIH application. At right is shown the student's ^{35}S-labeled autoradiogram, in which the band pattern was shadow-enhanced (green arrows). An artifactual lane distortion is denoted by the red arrows, which is weakly indicated in the photocopy.

clearly exposed a similar artifactual distortion of the lanes, as shown in Figure 7. The mentor had falsified the preparation, the experimental conditions, and the molecular weights in the photocopy that he had submitted to the NIH.

<u>Recovery of Probative Details</u>: Examinations of images may even reveal new evidence that bears upon other elements that are required for a finding of scientific misconduct. In another case, the allegation involved six instances where different sets of autoradiograms were allegedly falsely labeled and presented as different experiments. The student claimed these were honest errors, due, in part, to her inadvertent use of unlabeled autoradiograms. However, image enhancement by one of the institution's committee found evidence that the original label on one autoradiogram had been

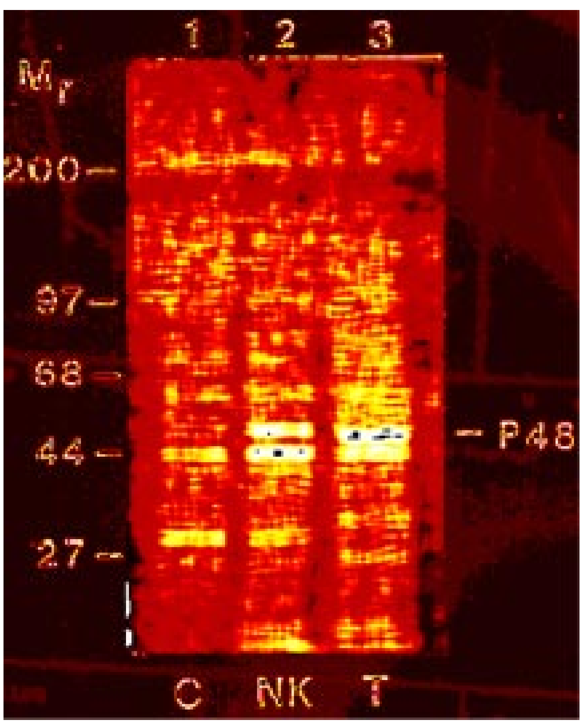

Figure 7. Computer enhancement of the bad photocopy shown in Figure 6. In ORI's image analysis, the distracting artifact in the photocopy can be removed by filtering, while false-coloring further enhanced the bands. The lane distortion artifact, present in the student's autoradiogram (Figure 6) was apparent in the same relation to the bands in the enhanced image, showing the student's results were falsified by the mentor to NIH.

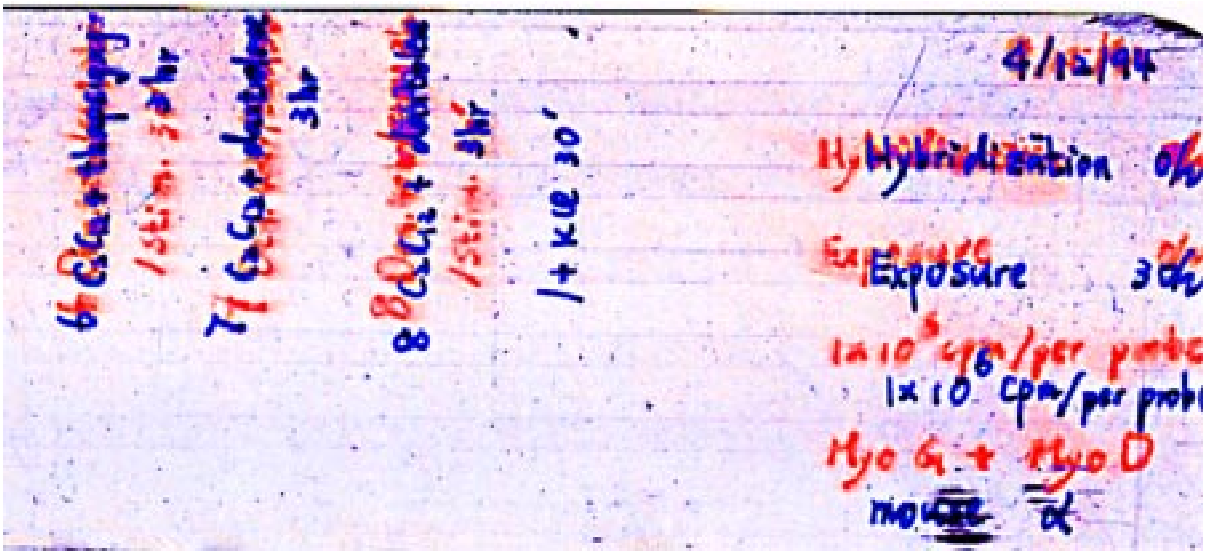

Figure 8. An example from one of six sets in which an autoradiogram had been falsely labeled and re-used. The institution's image analysis found evidence that the label for the prior experiment had been erased on the corner of the autoradiogram. The visible ink is blue, while evidence for the enhanced erasures is shown in red. Originally barely visible only as a faint and diffuse blue smear, the erased label differed from the film's background by only one gray level out of 256. The erasures were visualized here by ORI, after the film had been scanned at 42 bit resolution and the erasures had been selected for enhancement using their hue. The erased "MyoG" and "MyoD" denoted experiments on chickens and not mice. Thus, the re-use of the autoradiogram could not have been an honest error from mixup of unlabeled films, as the respondent originally claimed.

erased, but not fully (Figure 8). Thus, image processing revealed evidence that the falsification was not an honest error. ORI's subsequent analysis of figures in publications found that there was a pattern as to the six instances of re-use that was not consistent with their selection by chance.

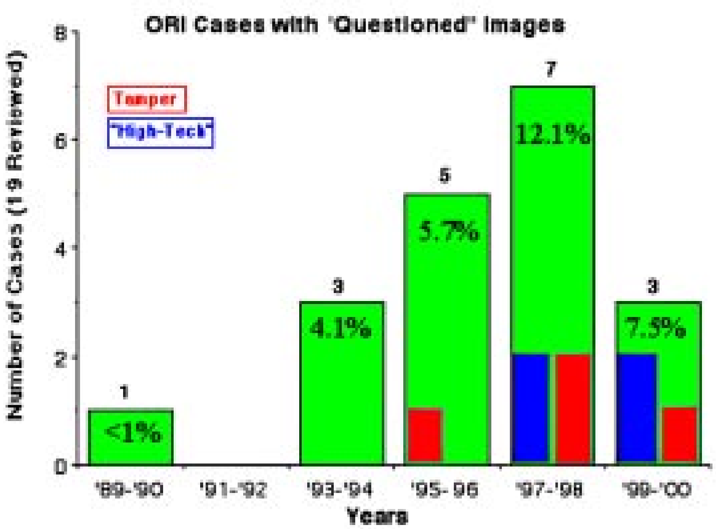

Figure 9. Incidence of 19 ORI cases involving contested scientific images. The data reflect when ORI's case file was opened; this formal step can occur at any phase in a case's history (i.e., at the allegation assessment, inquiry, or investigation stages). Thus the act of misconduct differs slightly in its date of occurrence. The percentages indicate the fraction of all ORI cases opened in those years. "Tamper" refers to allegations where the intensity of bands was selectively altered. "High-Tech" indicates manipulation by a computer to modify the image.

A scientific image is simply a picture constituting evidence that a test was carried out and/or that the test produced a certain outcome. In this context, the image is construed as qualitative "data." It could also be the basis for quantitative measurements, i.e., by measuring the "size" of a substance, or as the raw data for determine the amount of a substance. Thus, one consequence of discovering the falsification in an image is that there may be false claims elsewhere in a paper.

Compilation of Information from 19 ORI image analysis cases

In all of the cases above, the questioned image qualified as

Type of Misconduct Alleged	Number of ORI Cases
Falsely labeled as a different experiment (re-use)	13
Falsify molecular weight marker positions	>13
Cut, Graft, and Reuse, alter lane positions to fabricate data	5
Tampering: selective manipulation of image content, enhance/subtract bands, move position	4
Plagiarism of images (from Internet or journals), with false claims of element(s) from above	3

Table 1. Falsification of data with images–compilation from review of 19 ORI cases. This compilation indicates the incidence as number of cases, which under- represents the instances of misconduct, i.e., the actual number of figures or publications involved. The impact of the acts in each case was, in some cases, dramatically higher; one case involved 40 separate figures and multiple publications.

Image Source	**Respondent**
Thesis (student)	8 (5 students) (3 mentors)
Others (plagiarized)	3
Prior publication (self)	2
Status:	
Senior Faculty	7
Junior Faculty	4
Fellows	3
Students	5
Allegation Source:	
Student/Mentor/Co-Invest.	9
Reviewers	5
Inquiry Committee	2
Audiovisual Technician	1
Audience	1

Table 2. Characteristic of allegations of falsification of images in 19 ORI cases.

data, the intentional fabrication or falsification of which is the key element of research misconduct. On three occasions, a component of the allegation also involved plagiarism. The allegations reviewed by ORI generally involved use of fairly low-tech mechanisms for misrepresenting the data (Figure 9), such as re-use with false labels; in one case there were multiple instances of use of a felt-tip pen to add a band. Use of a computer to alter the content of the image has been seen less frequently.[6]

Table 1 compiles the nature of the misrepresentations involving questioned images in 19 ORI cases. The most common allegation was falsification of data by misrepresenting the results as being a different experiment, which also includes the attendant falsification of molecular weights. Only five examples occurred in which the lane markers were cut, re-grafted, and shifted so as to fabricate a test that was never run. Purposeful tampering with the image to selectively enhance or remove bands has occurred, but it was not very common. The allegations of plagiarism involved falsification of figures copied from published journal figures or by use (and falsification) of images obtained from the Internet homepages of other scientists.

Other aspects of these image-related allegations are described in Table 2. Thesis research appears to provide a relatively frequent source of questioned images, falsified by both students and mentors. In three cases, the images were allegedly obtained from others, and in two other cases they involved falsification of images that had been published earlier by the same laboratory. The source of most of these allegations was co-workers, although in five cases it was a reviewer who recognized the image as being from another source, or saw intrinsic evidence that the image could not be authentic. Most allegations did not arise because the images looked inauthentic, but simply because they were either recognized as false or represented claims that a reviewer frankly disbelieved. The questioned image was often the one concern in a case that could not be easily dismissed.

Discussion

The facts uncovered by the forensic examination of questioned images can often be corroborated by other evidence, such as absence of records/ experiments on the requisite date(s), the existence of dated computer files or other versions, parallel images in publications, etc. In addition to the basic image processing, a clear follow-up analysis is important.

The most useful analysis of questioned scientific images is done with a clear understanding of the experiment in question. This often requires collaboration between the image analysis and individuals who have a direct familiarity with the conduct of the scientific experiments at issue (9). To date, only two institutions have reported to ORI using a computer-based image analysis. Only one institution documented those results; in that instance, image processing by a committee member uncovered details that were determinative (see Figure 8). The information from ORI's cases indicates that most allegations involved "reuse" of the image to represent data from a purportedly different experiment. Occasionally, photographs of gels or blots were "cut and pasted" together in different combinations. Manipulations by computer were less common.

An image by itself creates a mantle of authenticity, if only because we give unusual weight to what we see. Yet in those cases where scientific misconduct was found, discovery of one falsified image often led to the discovery of another, and in all the "original' laboratory records were "missing." Thus good laboratory practice may help to deter or to minimize the impact of falsification.

Notes

1. Any views expressed in this article are my own and do not necessarily reflect those of the Office of Research Integrity. The citation of items in this article does not connote a product endorsement.
2. The questions are not limited to examining items that look *alike*. For example, immunoblots from the same gel can be stripped and re-probed with entirely new labeled antibody to reveal different protein bands.
3. The forensic value of the background information is completely analogous to the basis for numerical forensic analyses developed by Dr. James Mosimann in another presentation at this meeting.
4. A simple "thought" experiment makes the point more elegantly than examining the underlying physiology of visual perception: any two gray levels, so alike that they could be fairly represented as one shade, could still be assigned two separate colors, say red and blue, of the same intensity. (8)
5. Notice that digitizing at greater bit depth, such as 12 bit, would *in principle* detect fainter differences in shading to $^1/_{4096}$ parts, rather than the $^1/_{256}$ parts shown here.
6. It is debatable as to whether it would be more or less difficult to detect computer alterations. What can be said is that an allegation rarely arose because an image on its face appeared inauthentic.

Bibliography

1. Blitzer, Herbert, "Digital Imaging," Institute of Forensic Image Processing, Purdue School of Engineering and Technology, IUPUI. PDF document available on the Internet at http://www.engr.iupui.edu/ifi/articles. Published in Law Enforcement Technology, March 2000.
2. Scientific Working Group for Imaging Technologies, "Definitions and guidelines for the use of Imaging Technologies in the Criminal Justice System" (vs. 2.1, June 1999). Department of Justice PDF document available on the Internet at http://www.for-swg.org/swgitin.htm.
3. NIH Image is public domain software that was developed at the National Institutes of Health. A large number of analytical macros for NIH Image are available via an active electronic bulletin board. NIH Image is particularly useful for any quantitative and analytical measurements, false-coloring to better visualize or code details of processed images, but it is limited to 8 bit gray scale images on the Macintosh. NIH Image is available on the Internet at http://rsb.info.nih.gov/nih-image/.
4. Photoshop is a popular commercial program for both computer platforms (Windows and Macintosh). Photoshop records the history of the computer's actions, to document the steps used in the image processing, but its analytical capabilities are limited. Photoshop can be obtained from Adobe Systems, Inc., San Jose, CA 95110-2704.
5. Photoshop-compatible "plugins" are available commercially as the Image Processing Tool Kit® (IPTK), Reindeer Games, Inc., 20 Battery Park Ave., Suite 502, Asheville, NC 28801. IPTK is also compatible with NIH Image; it greatly extends the analytical capabilities of either program for analysis of 8 bit gray or 24 bit color images. Information can be found on the Internet at http://members.aol.com/ImagProcTK.
6. The base in support for analytical macros for ImageJ is not as large as the forerunner, NIH Image, but it is growing. ImageJ and its plugins can be obtained on the Internet at http://rsb.info.nih.gov/ij/plugins.
7. Inoué, Shinya, "Video Microscopy," Plenum Press, New York, N.Y. 1986. pp 80-83.
8. Russ, John C. "The Image Processing Handbook," CRC Press, Inc., Boca Raton, 2nd Edition. 1994. p. 213.
9. Indeed, the examples presented in this paper represent the collaboration and assistance from my ORI colleagues, Drs. John Dahlberg, Kay Fields, and Nancy Davidian.